

УДК: 004.94

Методы прогнозирования и модели распространения заболеваний

М. А. Кондратьев

Национальный исследовательский университет «Высшая школа экономики»,
лаборатория «Социология образования и науки»,
Россия, 190008, г. Санкт-Петербург, ул. Союза Печатников, д. 16

E-mail: mkondratyev@hse.ru

Получено 1 сентября 2013 г.

Число работ, посвященных прогнозированию инфекционной заболеваемости, стремительно растет по мере появления статистики, позволяющей провести анализ. В настоящей статье представлен обзор основных решений, доступных сегодня для формирования как краткосрочных, так и долгосрочных проекций заболеваемости; указаны их ограничения и возможности практического применения. Рассмотрены традиционные методы анализа временных рядов — регрессионные и авторегрессионные модели; подходы, опирающиеся на машинное обучение — байесовские сети и искусственные нейронные сети; рассуждения на основе прецедентов; техники, базирующиеся на решении задачи фильтрации. Перечислены важнейшие направления разработки математических моделей распространения заболевания: классические аналитические модели, детерминированные и стохастические, а также современные имитационные модели, сетевые и агентные.

Ключевые слова: прогнозирование заболеваемости, поточечные оценки, регрессионные модели, АРПСС, скрытые марковские модели, метод аналогий, экспоненциальное сглаживание, SIR, модель Барояна–Рвачева, клеточные автоматы, популяционные модели, агентные модели

Forecasting methods and models of disease spread

M. Kondratyev

*National Research University Higher School of Economics, Sociology of Education and Science Laboratory,
16 Ulitsa Soyuz Pechatnikov, St. Petersburg, 190008, Russia*

Abstract. — The number of papers addressing the forecasting of the infectious disease morbidity is rapidly growing due to accumulation of available statistical data. This article surveys the major approaches for the short-term and the long-term morbidity forecasting. Their limitations and the practical application possibilities are pointed out. The paper presents the conventional time series analysis methods — regression and autoregressive models; machine learning-based approaches — Bayesian networks and artificial neural networks; case-based reasoning; filtration-based techniques. The most known mathematical models of infectious diseases are mentioned: classical equation-based models (deterministic and stochastic), modern simulation models (network and agent-based).

Keywords: morbidity forecasting, point-to-point estimates, regression models, ARIMA, hidden Markov models, method of analogues, exponential smoothing, SIR, Rvachev–Baroyan model, cellular automata, population-based models, agent-based models

Citation: *Computer Research and Modeling*, 2013, vol. 5, no. 5, pp. 863–882 (Russian).

Введение

Методы прогнозирования инфекционной заболеваемости активно развиваются с начала XX века. В последние годы число работ на эту тему стремительно растет благодаря развертыванию информационных систем надзора и появлению больших объемов статистики, доступной для анализа. Эпидемиологические прогнозы выполняются для различных сроков и в зависимости от них служат разным целям. Так, краткосрочный прогноз на несколько недель вперед применяется в оперативном управлении и при выявлении эпидемических вспышек заболеваемости. Наиболее полезным можно считать среднесрочный прогноз сроком от двух месяцев до полугода [Myers et al., 2000], используемый в тактическом управлении. Конечно, он менее точен, нежели краткосрочный, но оставляет достаточно времени для подготовки к возможным чрезвычайным ситуациям и проведения превентивных мероприятий. При принятии стратегических решений не обойтись без долгосрочных прогнозов на год вперед и более. Достижение высокого качества таких прогнозов в большинстве случаев невозможно, тем не менее они требуются, например, при оценке необходимых объемов производства лекарственных препаратов и вакцин, оснащении медицинских учреждений и подготовке персонала.

Настоящая работа представляет собой обзор разработанных подходов к прогнозированию инфекционной заболеваемости и развития эпидемического процесса. Создано огромное количество техник прогнозирования, и едва ли возможно охватить все их вариации в одной статье. Поэтому в данной работе предпринята попытка выделить наиболее значимые и употребительные из предложенных методов и очертить область их применения.

Так, в зависимости от сроков прогнозирования и доступной статистики целесообразно использовать одни или другие подходы. Основу для анализа составляют временные ряды заболеваемости, которые могут дополняться данными различной природы — например, характеристиками погодных условий [Soebiyanto, Kiang, 2010]. Частота сбора данных обуславливается видом инфекции, текущей эпидемиологической обстановкой и организационными возможностями. В западных странах статистику заболеваемости стремятся обновлять ежедневно или чаще [Burkom, Murphy, Shmueli, 2007], в России — несколько реже. В частности, эпидемиологические данные по гриппу и острым респираторным вирусным инфекциям (ОРВИ) при отсутствии эпидемий собирают еженедельно [Система надзора за гриппом, 2010].

Все методы, рассматриваемые в настоящем обзоре, могут быть проиллюстрированы на примере прогнозирования заболеваемости гриппом. С одной стороны, грипп и ОРВИ занимают ведущее место среди инфекционных болезней человечества и, несмотря на все проводимые противоэпидемические мероприятия, заболеваемость ими не имеет тенденции к снижению как в России, так и за рубежом. Поэтому многие из моделей были предложены специально для анализа распространения гриппа. С другой стороны, техники прогнозирования гриппа подходят непосредственно или могут быть легко адаптированы для других инфекций, передающихся воздушно-капельным и контактно-бытовым путем. Впрочем, значительная часть методов применима для прогнозирования даже неинфекционной заболеваемости.

Для удобства изложения разобьем существующие способы прогнозирования заболеваемости на несколько категорий: статистические методы, методы на основе машинного обучения и прецедентов, методы на базе фильтрации и математическое моделирование распространения инфекции. Это весьма условное разделение, так как ряд подходов находится на стыке выбранных категорий, а некоторые техники даже могут быть представлены как частный случай других методов. Тем не менее все они имеют самостоятельное значение и особенности практического применения.

Статистические методы прогнозирования

Обозначим наблюдаемое значение показателя заболеваемости в момент времени t как y_t . Последовательность равноотстоящих значений y_t формирует временной ряд, который следует рассматривать как одну из возможных реализаций случайного процесса заболеваемости. В качестве примера на рисунке 1 изображено число зарегистрированных обращений за медицинской помощью лиц с симптомами гриппа и ОРВИ в Москве в 2000–2012 годах.

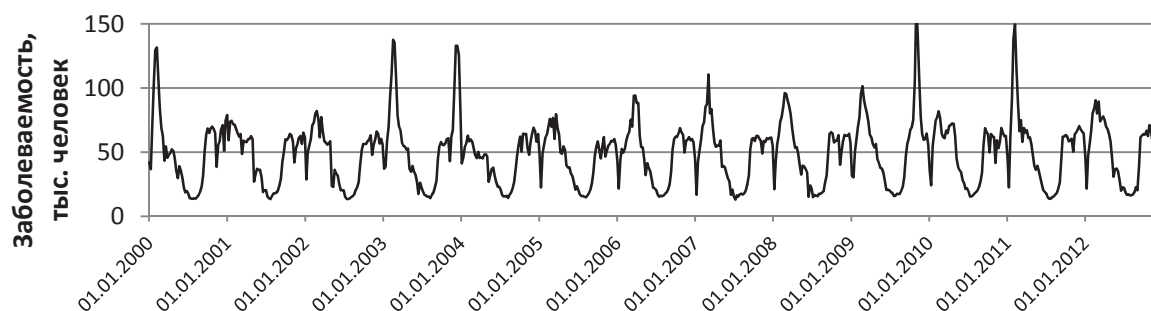


Рис. 1. Заболеваемость гриппом и ОРВИ в Москве по календарным неделям

Поточечные оценки

Для большинства инфекций, как и для ОРВИ, характерен сезонный подъем заболеваемости, и зачастую близкие значения показателей повторяются из года в год. Выраженная сезонность учитывается при любом прогнозе сроком более месяца. В простейшем случае данные за определенный календарный период (неделю, месяц года) рассматривают обособленно. Пусть T — период сезонности, тогда прогноз $\hat{y}_t, \hat{y}_{t+1}, \dots, \hat{y}_{t+T}$ может быть вычислен для каждого сечения процесса заболеваемости на основе множеств известных значений $\{y_{t-T}, y_{t-2T}, y_{t-3T}, \dots\}, \dots, \{y_{t-1}, y_{t-1-T}, y_{t-1-2T}, \dots\}$ соответственно.

Так, распространенный способ получения оценок ожидаемой заболеваемости [A practical guide for designing... 2008] — простое усреднение: $\hat{y}_t = \sum_{j=1}^n y_{t-jT}/n$, где n — количество доступных наблюдений. Этот подход широко применяется при вычислении уровня ординарной заболеваемости. Например, действующая в России методика расчета эпидемических порогов по гриппу и ОРВИ предполагает усреднение данных за календарные недели каждого года с одинаковым номером в течение 5–10 последних лет [Киселев и др., 2010]. Такой способ прогнозирования предполагает, что ежегодная заболеваемость носит неизменный характер и наблюдения в течение каждого нового года — следующая реализация одного и того же случайного процесса.

В подавляющем большинстве случаев это допущение не может быть соблюдено. Если в динамике заболеваемости присутствует отчетливый тренд, учесть его можно с помощью расширения подхода — поточечных линейных оценок (point-to-point linear estimates [Serfling, 1963]). Тогда наблюдения $\{y_{t-T}, y_{t-2T}, \dots\}$, соответствующие одному и тому же промежутку времени, не усредняются, а аппроксимируются прямой линией. При этом наклон линий для всех сечений в году выбирается общий, а свободный член в уравнениях прямых — индивидуальный. Это уместно, если дисперсионный анализ не выявляет существенных различий между наклонами регрессионных линий каждого из сечений. Несложно представить обобщение такой техники и для других видов регрессионных зависимостей между данными, хотя на практике необходимость в этом обычно отсутствует.

Процедуры прогнозирования, основанные на поточечных оценках, привлекают своей простотой. Несмотря на то что они не учитывают корреляцию между последовательными наблюде-

ниями, они могут успешно применяться для получения грубых среднесрочных и долгосрочных проекций заболеваемости.

Регрессионные модели

Классический регрессионный анализ — наверное, самый популярный метод прогнозирования заболеваемости. Задача регрессии заключается в нахождении оценок неизвестных параметров и формировании функциональной зависимости между заболеваемостью и факторами, ее предопределяющими. Среди регрессионных моделей следует выделить две группы [Burkom, Murphy, Shmueli, 2007]: неадаптивные модели, для оценки параметров которых используются все имеющиеся данные, и адаптивные модели, значения параметров которых рассчитываются на основе скользящего окна наблюдений.

Вид регрессионной зависимости для неадаптивной модели, конечно, выбирается исходя из свойств анализируемого временного ряда. В отдельных случаях можно применять полиномиальные или степенные функции [Кизим, Доровской, 2010], но чаще модель должна учитывать сезонный характер заболеваемости. Для этих целей используют предложенную Серфлингом циклическую функцию вида [Serfling, 1963; Pelat et al., 2007]

$$\hat{y}_t = \sum_{j=0}^{\nu} \alpha_j t^j + \sum_{j=1}^{\kappa} (\beta_{2j-1} \sin \theta_j + \beta_{2j} \cos \theta_j), \quad (1)$$

где \hat{y}_t — оценка заболеваемости в момент времени t , α_j и β_j — параметры регрессии, степень полинома ν обычно равна единице, а θ_j — линейная функция времени t . Ее следует выбирать исходя из результатов спектрального анализа [Sumi, Kato, 2012]. Обычно используют $\theta_j = 2\pi jt/T$, где T — период сезонности, например, 12 месяцев или 52 недели. Число гармоник κ редко превышает две.

Если доступна дневная статистика заболеваемости, модель Серфлинга дополняют индикаторными функциями $I_j(t)$, позволяющими учесть неравномерность обращений за медицинской помощью в течение недели [Brillman et al., 2005]:

$$\hat{y}_t = \sum_{j=0}^{\nu} \alpha_j t^j + \sum_{j=1}^{\kappa} (\beta_{2j-1} \sin \theta_j + \beta_{2j} \cos \theta_j) + \sum_{j=1}^7 \eta_j I_j(t). \quad (2)$$

Здесь η_j — параметры регрессии, а $I_j(t)$ равна единице, если t соответствует j -му дню недели и принимает нулевое значение в остальных случаях. Аналогично выделяют праздничные дни.

Неадаптивные модели такого вида предполагают, что сезонность каждый год проявляется схожим образом. Если это не так, возможно применение более сложных иерархических моделей [Brillman et al., 2005].

Модель Серфлинга — однофакторная и предназначена для экстраполяции исторически сложившихся эпидемических кривых с выраженной сезонностью. В том случае, если число наблюдений мало (менее пяти лет при периоде сезонности в один год), но доступны временные ряды нескольких предикторов, применяют пуассонову регрессию вида [A practical guide for designing. . . , 2008; Vergu et al., 2006]

$$\hat{y}_t = \gamma \exp \left\{ \sum_{j=0}^{\nu} \alpha_j t^j + \sum_{j=1}^{\kappa} (\beta_{2j-1} \sin \theta_j + \beta_{2j} \cos \theta_j) + \rho_1 x_t^1 + \rho_2 x_t^2 + \dots \right\}. \quad (3)$$

Здесь γ и ρ_j — параметры, а x_t^i — значения факторов, влияющих на заболеваемость, например, характеристик вируса, погодных условий или показателей продаж медикаментов. Также Пуассонову регрессию можно использовать при необходимости оценки суммарной заболеваемости

разными видами инфекций. Тогда в качестве x_t^i выступает число больных каждой инфекцией [Thompson et al., 2003], рассчитанное, например, по результатам лабораторной диагностики. При отсутствии сезонности применяют и полиномиальные многофакторные регрессионные модели [Честнова, Смольянинова, Логвинов, 2011].

Неадаптивные регрессионные модели призваны учитывать всю предысторию заболеваемости на анализируемой территории. Для их построения используются все имеющиеся данные или, по крайней мере, наблюдения последних лет, обладающие схожими характеристиками. Так, если свойства процесса заболеваемости изменились, вероятно, устаревшие данные уже не помогут уточнить прогноз. Адаптивные регрессионные модели представляют ограниченный отрезок временного ряда и должны быть достаточно чувствительны для того, чтобы реагировать на текущие тенденции заболеваемости.

Важную роль при использовании адаптивных моделей играет ширина скользящего окна — число последних наблюдений, на основе которых оценивают параметры модели. Ее следует выбирать исходя из срока прогнозирования. Как правило, адаптивные модели применяют при краткосрочном прогнозировании и ширина окна не превышает нескольких месяцев. Однако ширина окна тесно связана с видом регрессионной зависимости. Чем больше число неизвестных параметров, тем больше требуется данных для получения их оценок, и наоборот, чем шире окно, тем сложнее может потребоваться функция для адекватного представления участка эпидемической кривой. Например, при прогнозировании заболеваемости гриппом и ОРВИ на 1–2 недели вперед с помощью простой линейной регрессии [Cowling et al., 2006] скользящее окно должно вмещать данные 4–6 недель. В общем случае, ширина окна и даже вид регрессионной зависимости могут динамически меняться в течение года или в соответствии с характером эпидемической обстановки.

Так же, как и неадаптивные модели, при наличии подневной статистики заболеваемости адаптивные дополняют индикаторными функциями для представления соответствующих неравномерностей учета, например [Burkom, Murphy, Shmueli, 2007]:

$$\hat{y}_t = \sum_{j=0}^v \alpha_j t^j + \sum_{j=1}^7 \eta_j I_j(t) + \zeta I_{hol}(t). \quad (4)$$

Здесь α_j , η_j , ζ — параметры, $I_j(t)$ и $I_{hol}(t)$ — индикаторные функции дня недели и праздничного дня. При наличии необходимых данных целесообразно также применять адаптивные многофакторные модели [Boyle et al., 2011].

Регрессионный анализ — базовый подход к прогнозированию временных рядов любой природы, легко реализуемый с помощью любой системы компьютерной математики. Неадаптивные модели позволяют получить проекции заболеваемости на любой срок, но такие модели игнорируют локальные колебания эпидемических показателей и плохо подходят для краткосрочного прогнозирования. Адаптивные модели, напротив, предназначены для формирования прогнозов на несколько недель вперед. При достаточно большой ширине скользящего окна могут быть рассчитаны и среднесрочные оценки заболеваемости, но достижение высокого их качества проблематично.

Использование вышеописанных форм регрессии предполагает, что невязки модели (ошибки соответствия модели наблюдениям) — независимые случайные величины, имеющие одинаковый закон распределения с нулевым математическим ожиданием и постоянной дисперсией. Распространенной проблемой является наличие существенной автокорреляции невязок, которая обычно указывает на то, что модель может быть дополнена, а прогноз — уточнен, например, с помощью авторегрессионных моделей.

Авторегрессионные модели

Чаще всего авторегрессионные модели применяются в методологии Бокса и Дженкинса [Бокс, Дженкинс, 1974; A practical guide for designing... , 2008], в рамках которой измерения временного ряда представляются в виде

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}. \quad (5)$$

Здесь μ — «уровень» случайного процесса (константа, подлежащая определению), ϕ_j и θ_j — неизвестные параметры, а ε_j — случайные импульсы невязок модели (распределенные нормально с нулевым математическим ожиданием и постоянной дисперсией σ_ε^2 , которую оценивают по наблюдениям). Приведенное выражение называют смешанной моделью авторегрессии — скользящего среднего (АРСС) порядка (p, q) . АРСС описывают только стационарные временные ряды. В редких случаях возможно применять АРСС для краткосрочного прогнозирования заболеваемости [Lai, 2005; Вьон и др., 2011], но чаще требуется преобразование анализируемого временного ряда.

Общепринятым для этих целей является взятие разности $y_t - y_{t-1} = (1 - L)y_t$, где L обозначает разностный оператор $Ly_t = y_{t-1}$. С его помощью можно переписать выражение (5) как

$$\left(1 - \sum_{j=1}^p \phi_j L^j\right) (1 - L)^d y_t = \mu + \left(1 - \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t, \quad (6)$$

где d — порядок разности, который требуется для достижения стационарности (обычно не превышает двух). Такая модель представляет процесс авторегрессии — проинтегрированного скользящего среднего (АРПСС) порядка (p, d, q) .

Наконец, при исследовании сезонных процессов с периодом T измерения временного ряда связывают не только с последними наблюдениями, но и с наблюдениями, выполненными T моментов времени назад. Для этого модель дополняют соответствующими множителями, аналогичными по смыслу уже имеющимся:

$$\left(1 - \sum_{j=1}^p \phi_j L^j\right) \left(1 - \sum_{j=1}^P \Phi_j L^{Tj}\right) (1 - L)^d (1 - L^T)^D y_t = \mu + \left(1 - \sum_{j=1}^q \theta_j L^j\right) \left(1 - \sum_{j=1}^Q \Theta_j L^{Tj}\right) \varepsilon_t. \quad (7)$$

Идентификация определения порядка $(p, d, q)(P, D, Q)_T$ — сезонной (мультипликативной) модели АРПСС (7) на практике бывает достаточно сложна и неоднозначна [Бокс, Дженкинс, 1974] и требует тщательного анализа временного ряда. Предложены автоматические процедуры идентификации моделей [Hundman, Khandakar, 2008], которые значительно упрощают применение АРПСС, но не всегда позволяют достичь оптимальных результатов.

Тем не менее при прогнозировании заболеваемости АРПСС немногим уступает по популярности регрессионным моделям [Гальченко и др., 2010; Soebiyanto, Adimi, Kiang, 2010]. При необходимости расширенные модели АРПСС позволяют учитывать влияние внешних факторов [Chen et al., 2011; Díaz-Hierro et al., 2012]. АРПСС — метод, используемый в Центре по контролю за заболеваемостью (США, Атланта) с 1980-х годов [Choi, Thacker, 1982], который пришел на смену модели Серфлинга. АРПСС подходит как для краткосрочного, так и для долгосрочного прогнозирования, алгоритмы оценивания параметров модели реализованы во всех специализированных пакетах статистических программ.

Многие временные ряды не могут строго удовлетворить требованию стационарности, что влечет за собой неточности при применении АРСС. Кроме того, для оценки параметров

АРПСС необходимо большое количество наблюдений. При малом количестве исходных данных предпочтительно использование иной техники — целочисленной авторегрессии (INAR — integer autoregressive model) [Moriña et al., 2011; Unkel et al., 2012], которую можно отнести к активно развиваемым способам прогнозирования. Существуют и другие варианты авторегрессионных моделей, например, обобщенные модели авторегрессии — скользящего среднего (GARMA — generalized autoregressive moving average model) [Dugas et al., 2013].

Прогнозирование на основе машинного обучения и прецедентов

Байесовские сети

Другой подход к моделированию временных рядов, обеспечивающий учет структуры корреляции данных, — динамические байесовские сети [Sebastiani et al., 2006; Unkel et al., 2012]. Байесовские сети изображают в форме направленного графа, вершины которого соответствуют переменным модели, а ребра — вероятностным зависимостям между ними, которые заданы определенными законами распределения. После обучения байесовские сети позволяют оценить вероятность наступления некоторого события при наблюдаемой последовательности явлений. Байесовские сети сегодня стремительно набирают популярность в различных областях знаний и применяются к задаче прогнозирования заболеваемости преимущественно в простой форме скрытых марковских моделей (СММ).

Основной идеей СММ является сопоставление каждой случайной величины Y_t с ненаблюдаемой случайной величиной S_t , определяющей условное распределение Y_t [Le Strat, Carrat, 1999; Siettos, Russo, 2013]. Задавая вид закона распределения, связывающего Y_t и S_t , можно оценить его параметры по наблюдаемым значениям y_t . Так, в качестве Y_t может выступать число обращений горожан за медицинской помощью, а под S_t подразумеваться важная характеристика эпидемиологической обстановки, например, общее число инфицированных горожан. СММ предполагают, что величина Y_t зависит только от значения скрытой переменной S_t в момент времени t , а последовательность S_t обладает марковским свойством, то есть величина S_t зависит только от S_{t-1} (рис. 2а).

Построение СММ возможно как при большом, так и малом количестве исходных данных, однако алгоритмы оценки параметров модели вычислительно сложны, поэтому СММ обычно анализируют на основе узкого скользящего окна наблюдений [Watkins et al., 2009]. По этой причине на текущий момент байесовские сети предоставляют возможность только краткосрочного прогнозирования заболеваемости. Более того, чаще всего СММ применяются лишь для выявления повышенной заболеваемости. При этом строится простая модель с двумя состояниями, соответствующими ординарной и эпидемической ситуации [Watkins et al., 2009; Unkel et al., 2012]. Разработка и анализ байесовских сетей и СММ выполняются с использованием специализированного программного обеспечения.

Искусственные нейронные сети

Важный представитель методов, опирающихся на машинное обучение, — искусственные нейронные сети (ИНС). Методология ИНС широко известна и хорошо подходит для решения задач, аналитическое исследование которых затруднительно. ИНС представляет собой направленный взвешенный граф, вершины которого моделируют функционирование биологических нейронов (рис. 2б). Вершины принимают входящие сигналы и при достаточно большом значении их взвешенной суммы преобразуют их в исходящий сигнал. Обучение ИНС заключается в вычислении коэффициентов связей между вершинами, определяющих силу входящих сигналов, и выполняется на основе эмпирических данных: статистики заболеваемости и при наличии, значений факторов, ее предопределяющих.

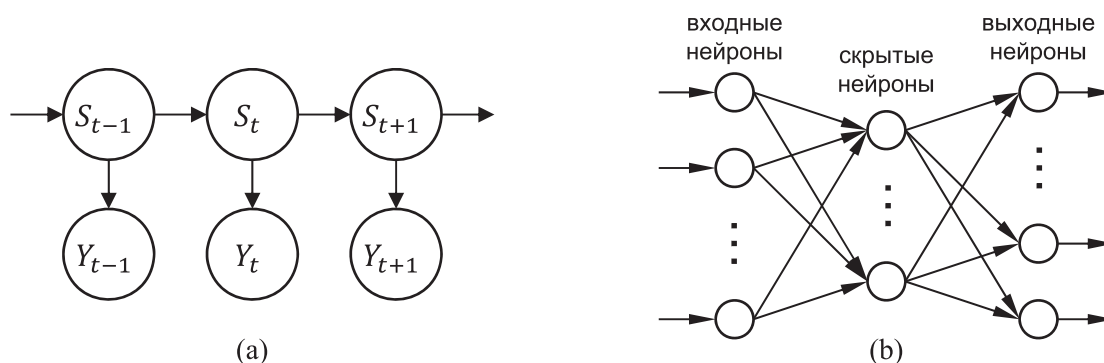


Рис. 2. Схема зависимостей в скрытой марковской модели (а); схема ИНС с одним скрытым слоем (б)

Допустимо множество вариантов применения ИНС для прогнозирования заболеваемости: ИНС могут отличаться архитектурой, количеством нейронов и видом функции их активации, способом обучения. Формального подхода к выбору характеристик ИНС не существует, в подавляющем большинстве случаев используются сети прямого распространения сигнала и алгоритм обратного распространения ошибки [Bai, Jin, 2005; Kiang et al., 2006].

Цель обучения ИНС — определение явных и неявных зависимостей между входными и выходными данными. Показатели заболеваемости, как правило, зашумлены, поэтому для обучения ИНС требуется сравнительно большое их количество. Достаточную для долгосрочного прогнозирования предысторию заболеваемости еще только предстоит накопить, поэтому пока ИНС подходят скорее для прогнозирования на малые сроки.

В целом же ИНС — весьма перспективный подход, распространенный не только в зарубежных, но и отечественных работах [Фирсов, 2006; Белецкая, Коровин, Родионов, 2010; Ефимова, Горнов, Зароднюк, 2010]. Методология ИНС реализована в большинстве систем компьютерной математики и статистических пакетов.

Рассуждения на основе прецедентов

Рассуждения на основе прецедентов (case-based reasoning) — семейство методов, которые используются не только при прогнозировании. Они базируются на идее поиска ответа на поставленную задачу среди уже известных способов ее решения.

Применительно к задаче прогнозирования заболеваемости этот подход называют методом аналогий [Viboud et al., 2003; Schmidt, Waligora, 2007]. Метод аналогий изначально был разработан для построения прогноза погоды. Он сводится к поиску среди накопленной истории наблюдений эпидемических ситуаций, близких к сложившейся к настоящему времени, и построению прогноза, исходя из последствий этих ситуаций.

Ключевым аспектом метода является вопрос о том, какие две ситуации являются похожими. Пусть y_t — заболеваемость в текущий момент времени t , тогда ситуацию можно охарактеризовать вектором наблюдений $\mathbf{y}_t = (y_t, y_{t-1}, \dots, y_{t-k})^T$. Сравнить настоящую заболеваемость \mathbf{y}_t с определенным моментом в прошлом \mathbf{y}_τ можно, например, с помощью метрики $dist(\mathbf{y}_t, \mathbf{y}_\tau) = \sum_{j=0}^k (y_{t-j} - y_{\tau-j})^2$. Конечно, и число анализируемых наблюдений k и сама метрика должны выбираться исходя из свойств временного ряда и срока прогнозирования.

Прогноз может быть сформирован по-разному. Например, отобрав n ситуаций $\{\mathbf{y}_{\tau_1}, \dots, \mathbf{y}_{\tau_n}\}$, аналогичных текущей \mathbf{y}_t , прогноз рассчитывают как $\hat{y}_{t+h} = \sum_{j=1}^n w_j y_{\tau_j+h}$, где веса w_j назначают исходя из близости ситуаций \mathbf{y}_t и \mathbf{y}_{τ_j} , $\sum_{j=1}^n w_j = 1$.

Метод аналогий рассчитан, прежде всего, на краткосрочное прогнозирование и не предполагает учет влияния на заболеваемость внешних факторов. Использование его для средних и больших сроков прогнозирования потребовало бы огромных массивов статистики, а также более сложной метрики, учитывающей как минимум эффекты старения информации. Но даже при их наличии поиск действительно близких ситуаций будет затруднителен, а любые различия будут отражаться на точности проекций [Dugas et al., 2013]. В отличие от всех ранее перечисленных техник, в литературе метод аналогий встречается сравнительно редко. Тем не менее это простая и эффективная методика, возможность применения которой следует учитывать.

Прогнозирование на базе фильтрации

Любые временные ряды заболеваемости можно рассматривать как случайный процесс, состоящий из сигнала, отражающего реальную эпидемическую обстановку, и высокочастотного шума. Фильтрация шума позволяет уточнить прогноз и может выполняться как в ходе предварительной обработки исходных данных, так и в составе самого алгоритма прогнозирования.

Вейвлет-декомпозиция

Отфильтровать измерения можно многими способами, один из них — вейвлет-декомпозиция. В ее основе лежит разложение временного ряда с помощью вейвлет-функций ϕ и ψ [Shmueli, Fienberg, 2006]:

$$y_t = \sum_{k=1}^n a_k \phi(t-k) + \sum_{k=1}^n \sum_{j=1}^m d_{j,k} \psi(2^j t - k), \quad (8)$$

где a_k — отмасштабированный сигнал в момент времени k , $d_{j,k}$ — детализирующий коэффициент, n и m — целые числа. Фильтрация достигается за счет ограничения множества значений коэффициентов разложения. Вейвлет-декомпозиция достаточно сложна вычислительно и в интерпретации, но в некоторых работах успешно используется при подготовке данных для прогнозирования заболеваемости [Гальченко и др., 2010; Unkel et al., 2012].

Вейвлет-декомпозиция обычно применяется совместно с другими техниками прогнозирования, но существуют и самостоятельные подходы, непосредственно базирующиеся на идее фильтрации.

Экспоненциальное сглаживание

К таким подходам принадлежит экспоненциальное сглаживание, представляющее собой частный случай взвешенного скользящего среднего. При экспоненциальном сглаживании значение заболеваемости y_t в момент времени t описывается взвешенной суммой последних наблюдений: $l_t = \alpha y_t + (1 - \alpha)l_{t-1}$. Здесь α — коэффициент сглаживания ($0 < \alpha < 1$), который обеспечивает уменьшение веса по мере старения данных, которое может рассматриваться как отражение естественного процесса обучения. Прогноз при этом рассчитывается как $\hat{y}_{t+h} = l_t$, где l_t — сглаженное значение.

В некоторых случаях простое экспоненциальное сглаживание применяют при прогнозировании заболеваемости [Головин, Петросян, 2009; Иванов и др., 2012], но оно не подходит для рядов, в поведении которых присутствуют отчетливый тренд или сезонность. Для этих целей используются обобщенные модели [Gardner, 1985; Hyndman, Khandakar, 2008].

В частности, заболеваемость гриппом и ОРВИ можно описать аддитивной сезонной моделью Хольта–Винтерса:

$$\begin{aligned}l_t &= \alpha(y_t - s_{t-T}) + (1 - \alpha)(l_{t-1} + r_{t-1}), \\r_t &= \gamma(l_t - l_{t-1}) + (1 - \gamma)r_{t-1}, \\s_t &= \delta(y_t - l_t) + (1 - \delta)s_{t-T}, \\ \hat{y}_{t+h} &= l_t + hr_t + s_{t-T+h},\end{aligned}\tag{9}$$

где r_t представляет тренд, s_t — сезонный фактор, γ и δ — коэффициенты сглаживания для них, а T — период сезонности. В аддитивной модели сезонные флуктуации не зависят от величины показателей заболеваемости, такая связь учитывается в мультипликативных моделях. Все эти модели — однофакторные. Результаты сглаживания в значительной степени определяются начальными значениями l_0 , r_0 , s_0 . Их выбор может осуществляться по-разному, общие рекомендации приведены в [Chatfield, Yar, 1988].

Прогностические методы, опирающиеся на экспоненциальное сглаживание, хорошо известны и очень популярны в экономике, реализованы во всех крупных статистических пакетах программ. Обобщенные модели пригодны как для краткосрочного, так и для долгосрочного прогнозирования уровня заболеваемости, но, почему-то, редко используются в работах по этой тематике [Burkom, Murphy, Shmueli, 2007; Díaz-Hierro, 2012].

Калмановская фильтрация

Класс моделей экспоненциального сглаживания пересекается, но не совпадает с классом рассмотренных ранее моделей АРПСС [Chatfield, Yar, 1988; Hyndman, Khandakar, 2008]. И те, и другие имеют эквивалентное представление в пространстве состояний. Так, любые эпидемические процессы можно описать следующей системой разностных уравнений:

$$\begin{aligned}\mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{w}_t, \\ \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t + \mathbf{D}\mathbf{f}_t + \mathbf{v}_t,\end{aligned}\tag{10}$$

где \mathbf{x}_t — вектор переменных состояния системы в момент времени t , \mathbf{y}_t — вектор наблюдений, \mathbf{f}_t — вектор значений внешних факторов, \mathbf{w}_t и \mathbf{v}_t — белый шум. Матрицы параметров \mathbf{A} , \mathbf{H} , \mathbf{D} определяют модель эпидемического процесса и выбираются исходя из решаемой задачи — краткосрочного или долгосрочного прогнозирования.

Такая форма записи позволяет предложить обобщенные модели распространения заболевания. Для формирования прогноза могут применяться методы теории регулирования, в частности калмановская фильтрация [Hamilton, 1994]. Фильтры Калмана широко используются в инженерных и эконометрических приложениях, но, за редким исключением [Shaman, Karspeck, 2012] пока не применяются в области эпидемиологического надзора. Впрочем, в литературе можно обнаружить примеры употребления и других фильтров, например согласованных [Burkom, 2003].

Математическое моделирование распространения инфекции

Все вышерассмотренные техники можно отнести к общим методам анализа временных рядов. Они не учитывают специфические свойства, присущие эпидемическим процессам, и опираются только на значения наблюдаемых показателей уровня заболеваемости. В том или ином виде они могут применяться при прогнозировании данных любой природы. Другая группа методов, созданных в рамках так называемого биологического подхода [Muys et al., 2000], основывается на особенностях механизма передачи инфекции и явно моделирует его. Разработано множество математических моделей распространения заболевания (далее МРЗ) такого типа, которые заслуживают отдельного обзора. Обозначим в настоящей статье лишь важнейшие направления работ в этой области.

Аналитические модели распространения заболевания

Классические МРЗ описывают динамику распространения инфекции с помощью систем дифференциальных уравнений. Первые работы, заложившие фундамент современной математической эпидемиологии, были опубликованы в начале XX века [Siettos, Russo, 2013]. Наиболее известной является модель SIR, предложенная в 1920-х годах Кермаком (Kermack) и МакКендриком (McKendrick), а также ее расширения (compartmental models). В таких моделях исследуемую популяцию разбивают на группы, например, на здоровых (S – susceptible), больных (I – infectious) и выздоровевших людей (R – recovered). Модели типа SIR опираются на аналогию между контактами людей в большой популяции и законом действующих масс в химической кинетике. Так, приращение числа людей в каждой из групп представляют с помощью системы уравнений вида

$$\frac{dS(t)}{dt} = -\alpha S(t)I(t), \quad \frac{dI(t)}{dt} = \alpha S(t)I(t) - \beta I(t), \quad \frac{dR(t)}{dt} = \beta I(t), \tag{11}$$

где α отражает вероятность передачи инфекции, а $1/\beta$ – длительность заразительного периода. В зависимости от решаемой задачи и особенностей инфекции могут выделяться различные категории людей [Coburn, Wagner, Blower, 2009] (рис. 3) – зачастую учитывают индивидов с заболеванием в инкубационном периоде (E – exposed), вакцинированных индивидов (V – vaccinated), возрастные группы населения. Например, в работе [Eichner et al., 2007] для моделирования демографической структуры региона используется более 1000 уравнений.

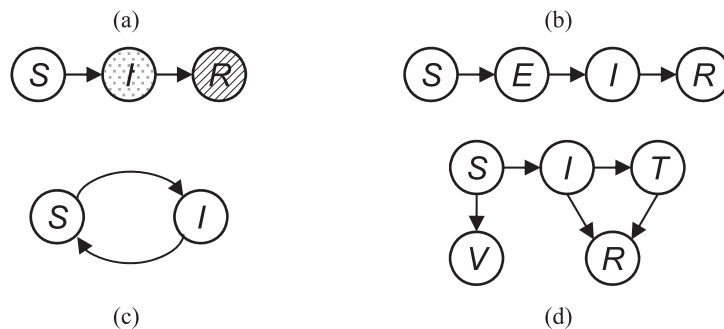


Рис. 3. Схема модели SIR (a); схема модели SEIR (b); схема модели SIS (c); схема модели типа SIR с учетом вакцинированных индивидов и индивидов, находящихся на лечении (T – treatment) (d)

Применение моделей SIR ограничивает допущение о достаточно большой численности анализируемой популяции, а также о том, что все индивиды в ней непрерывно и равномерно перемешиваются [Hethcote, 2000]. Разработано множество МРЗ, призванных детализировать SIR и сделать возможным их практическое использование. Например, в работе [Бао, 2012] для прогнозирования распространения ВИЧ применяется модификация SIR с динамическими параметрами (α, β). Наиболее значимая отечественная МРЗ – модель Барояна–Рвачева [Боев, Макаров, 2005], разработанная в СССР в 1960-х годах. Она описывает эпидемический процесс с помощью системы интегро-дифференциальных уравнений

$$\begin{aligned} \frac{dS(t)}{dt} &= -\frac{\lambda}{p(t)} S(t) \int I(t, \tau) d\tau, & \frac{\partial E(t, \tau)}{\partial t} + \frac{\partial E(t, \tau)}{\partial \tau} &= -\gamma(\tau) E(t, \tau), \\ \frac{\partial I(t, \tau)}{\partial t} + \frac{\partial I(t, \tau)}{\partial \tau} &= \gamma(\tau) E(t, \tau) - \delta(\tau) I(t, \tau), & \frac{dR(t)}{dt} &= \int \delta(\tau) I(t, \tau) d\tau, \end{aligned} \tag{12}$$

где p — общая численность популяции, λ , γ , δ — функции развития инфекционного процесса, t — глобальное, календарное время, а τ — локальное время, идущее с момента заражения индивида. Модель Барояна–Рвачева успешно применялась для прогнозирования эпидемий гриппа на территории СССР.

Детерминированные модели SIR критикуют за их неспособность отражать вероятностную природу процессов заболеваемости, поэтому были созданы стохастические варианты таких МРЗ [Daley, Gani, 1999]. Например, в работе [Colizza et al., 2007] рассматривается глобальная модель распространения атипичной пневмонии по территории 220 стран, основанная на модификации SIR. Авторы статьи [Бородулин и др., 2007] предложили статистическую версию модели Барояна–Рвачева.

Важным вопросом при моделировании эпидемий являются пространственные характеристики распределения инфекции и влияние географических факторов. К их учету подходят по-разному, например, в рамках методологии SIR можно представить различные географические точки, введя дополнительное деление на подгруппы. Тогда в систему уравнений включают механизмы перемещения индивидов между территориями с заданной интенсивностью [Araz et al., 2009]. Также для этих целей могут применяться гравитационные модели [Li et al., 2011] или их элементы в составе МРЗ типа SIR [Rinaldo et al., 2012].

Имитационные модели распространения заболевания

Значимым преимуществом моделей, базирующихся на аппарате дифференциальных уравнений, является возможность их аналитического исследования. Тем не менее для всех таких моделей характерно еще одно допущение — характеристики и поведение всех индивидов, отнесенных к одной подгруппе, считаются одинаковыми. Его позволяют ослабить современные компьютерные имитационные МРЗ.

Первыми имитационными МРЗ были клеточные автоматы [Бейли, 1970], которые представляют собой совокупность ячеек, каждая из которых принимает состояние из конечного множества. Обычно используются квадратные ячейки, объединенные в прямоугольную решетку. Тогда узлы решетки моделируют индивидов, каждый из которых имеет фиксированное положение в пространстве (рис. 4а). Время для клеточного автомата движется дискретно, на каждом шаге инфекция может с определенной вероятностью перекинуться от больного индивида к его восприимчивому соседу. В рамках такого подхода можно представить и перемещения индивидов [Sirakoulis, Karafyllidis, Thanailakis, 2000], но все же клеточные автоматы недостаточно реалистичны для практического применения.

Основным ограничением клеточного автомата является моделирование окружающего индивидов пространства как решетки. В действительности для большинства заболеваний можно считать, что инфекция распространяется посредством контактов между людьми. Однако контакты между индивидами не ограничиваются непосредственными соседями, они носят нерегулярный во времени и пространстве характер. Учесть их позволяют сетевые модели [Saramäki, Kaski, 2005], которые представляют схему возможных контактов в виде графа, динамического или статического (рис. 4б). Вершины графа здесь — объекты с набором индивидуальных свойств, сколь угодно детализированно описывающие состояние отдельных индивидов. Сетевые модели могут быть как детерминированными, так и стохастическими. МРЗ, основанные на клеточных автоматах, можно считать частным случаем сетевых МРЗ.

Сетевые модели предлагают общую технологию моделирования окружающего пространства, но не специфицируют способ организации графа контактов для населения конкретной территории. Наиболее естественным является формирование связей между индивидами и возможных путей распространения инфекции на основе их социальной принадлежности. Такой способ применяется в популяционных (population-based) МРЗ [Patel, Longini, Halloran, 2005; Chao et al.,

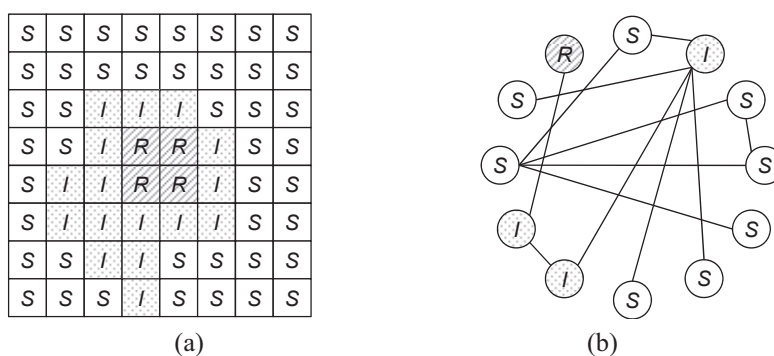


Рис. 4. Представление распространения инфекции клеточным автоматом (a) и сетевой моделью (b)

2010]. Окружающее пространство популяционных моделей разбито на локации — места, в которых индивиды могут контактировать, например, школы, офисы, жилые помещения и проч. В зависимости от возрастной и социальной группы индивида он связан с теми или иными локациями — ежедневно «посещает» их. Время в популяционных моделях движется дискретно, обычно с шагом в 12 или 24 часа. На каждом шаге рассчитывается вероятность заражения каждого индивида по формулам, учитывающим число инфицированных в посещаемых индивидом локациях. Популяционные МРЗ позволяют представить структуру общества любой сложности, но требуют тщательной настройки параметров на основе эмпирических данных.

Наиболее детальные и перспективные сегодня — агентные (agent-based) МРЗ [Das, Savachkin, Zhu, 2008; Ohkusa, Sugawara, 2009]. Они сохраняют подход популяционных МРЗ к моделированию окружающего пространства, но не используют централизованный расчет вероятностей передачи инфекции, а описывают поведение каждого человека — агента индивидуально. Большинство агентных МРЗ дискретно-событийные, то есть процесс их моделирования сводится к последовательной обработке списка запланированных событий. Каждое событие влечет за собой некоторый набор атомарных действий, меняющих состояние системы. Агенты самостоятельно формируют события, связанные с их деятельностью, что позволяет избежать фиксированного шага дискретизации и сложных формул вероятности инфицирования. Агентные МРЗ — максимально реалистичные, но требуют существенных вычислительных мощностей. Их калибровка и анализ, необходимые для использования модели в качестве инструмента прогнозирования, базируются на проведении множества численных экспериментов.

МРЗ позволяют сформировать прогноз на любой срок, но их исследование и практическое применение значительно сложнее других техник. МРЗ разрабатываются и настраиваются для решения конкретных задач, особенности которых отражают в самой модели и значениях ее параметров. Чаще всего МРЗ используют для исследования хода эпидемий и их последствий, поэтому такие модели плохо отражают ординарные показатели заболеваемости. Почти все МРЗ ориентированы на анализ одной единственной инфекции, механизм передачи которой заложен в структуру модели, и не позволяют прогнозировать объединенные показатели заболеваемости, например, по гриппу и ОРВИ. Для реализации МРЗ предпочтительно применять специализированные программные средства.

Смешанные техники прогнозирования

Рассмотренные методы используются как в чистом виде, так и в форме многочисленных модификаций и комбинаций. Например, параметры модели Серффлинга (1) определяют не только с помощью метода наименьших квадратов, но и робастной процедуры взвешивания, не требующей постоянной дисперсии невязок [A practical guide for designing... , 2008].

Распространенным приемом при прогнозировании временных рядов является декомпозиция. Показатель заболеваемости может быть представлен как $y_t = u_t + \gamma_t + c_t + \varepsilon_t$, где u_t — плавно меняющаяся компонента (тренд), γ_t — сезонная компонента, c_t — циклическая компонента, отражающая явления, повторяющиеся в течение длительного периода свыше одного года, ε_t — случайная компонента, формирующаяся под влиянием не поддающихся учету и регистрации случайных факторов. Декомпозиция — попытка отделения этих компонент для описания их моделями различной природы.

Например, сезонную составляющую ряда заболеваемости иногда выделяют простым точечным усреднением наблюдений с одинаковым периодом [Abeku et al., 2002; Permanasari, Rambl, Dominic, 2010]. Во всех статистических пакетах программ для этого реализован классический прием [Hyndman, Athanasopoulos, 2012], основанный на усреднении сглаженных значений временного ряда. Регрессионный анализ также может использоваться для декомпозиции. К примеру, можно описать медленно меняющиеся компоненты ряда (u_t , γ_t , c_t) регрессионной функцией, а оставшуюся изменчивость моделировать с помощью авторегрессионных техник [Lai, 2005]. Так, АРПСС хорошо подходит для прогнозирования невязок, если наблюдается их существенная корреляция [Reis, Mandl, 2003]. Такой подход можно рекомендовать для уточнения краткосрочных прогнозов, выполняемых на основе неадаптивных регрессионных моделей.

Фильтрацию шума также можно рассматривать как способ декомпозиции, позволяющий полностью или частично отбросить случайную компоненту ε_t . Так, вейвлет-декомпозицию применяют совместно с авторегрессионными методами [Гальченко и др., 2010]; аналитические МРЗ класса SIR объединяют с калмановскими фильтрами [Shaman, Karspeck, 2012]. Вообще, в области МРЗ, в частности имитационных МРЗ, возможности для смешения подходов к моделированию особенно широки.

Сочетания техник прогнозирования многообразны и используются авторами на свое усмотрение в зависимости от решаемой задачи. В общем случае, применение смешанных техник прогнозирования является уместным, если другие подходы не позволяют достичь требуемой точности прогноза.

Заключение

В таблице 1 перечислены основные методы прогнозирования инфекционной заболеваемости, рассмотренные в статье, и указаны предпочтительные сроки построения проекций. С алгоритмической точки зрения между среднесрочными и долгосрочными прогнозами нет различий, поэтому для их вычисления могут применяться одни и те же подходы. Следует заметить, что понятие срока прогнозирования в некоторой степени условно и зависит от шага дискретизации временного ряда. Так, например, если имеется последовательность ежегодных показателей заболеваемости, прогноз на год вперед можно отнести к краткосрочному. В настоящем обзоре предполагалось, что шаг дискретизации составляет не более одного месяца.

Как видно из таблицы, число методов, ориентированных на малые сроки прогнозирования, немногим больше числа пригодных для расчета проекций на длительные сроки. Но количество опубликованных работ распределено между ними отнюдь не пропорционально. Обсуждается множество способов вычисления и применения краткосрочных прогнозов, но долгосрочному прогнозированию посвящены лишь единичные статьи [Гальченко и др., 2010; Chen et al., 2011; Sumi, Kato, 2012]. Возможно, это связано с тем, что качество краткосрочных проекций существенно выше долгосрочных, а возможности их уточнения шире. Тем не менее, практическая ценность адекватных среднесрочных и долгосрочных прогнозов не меньше, а для многих задач и больше, чем краткосрочных.

Таблица 1. Методы прогнозирования инфекционной заболеваемости

	Краткосрочное прогнозирование	Среднесрочное и долгосрочное прогнозирование
Поточечные оценки		✓
Неадаптивные регрессионные модели		✓
Адаптивные регрессионные модели	✓	
Модели авторегрессии — проинтегрированного скользящего среднего	✓	✓
Динамические байесовские сети	✓	
Искусственные нейронные сети	✓	
Рассуждения на основе прецедентов	✓	
Обобщенное экспоненциальное сглаживание	✓	✓
Калмановская фильтрация	✓	✓
Математические модели распространения заболеваний	✓	✓

Все методы, указанные в таблице, за исключением математических МРЗ, могут использоваться при анализе и прогнозировании показателей инфекционной заболеваемости, неинфекционной заболеваемости и даже смертности. Более того, они подойдут почти для любых временных рядов биологического характера и не только. Главное назначение математических МРЗ — аналитическое или численное исследование эпидемических процессов. Их применение как средства прогнозирования ограничено моделированием распространения одноклеточной инфекции и его последствий.

Реализация некоторых из описанных подходов требует существенно больших затрат, чем остальных. Самый простой способ прогнозирования — поточечные оценки. Для их вычисления, конечно, можно воспользоваться СКМ, но и применение языка программирования высокого уровня (далее ЯП) здесь не вызовет затруднений. Немного сложнее будет построить проекции с помощью регрессионных моделей, рассуждений на основе прецедентов или обобщенного экспоненциального сглаживания. Эти методы интуитивно понятны и хорошо известны, для их успешного применения необходимо лишь уделить достаточное внимание выбору вида модельных зависимостей и параметров. Регрессионный анализ удобнее выполнять в СКМ (MATLAB, Maple, Mathcad и проч.). Метод аналогий допустимо реализовывать на ЯП — использование СКМ тут не даст преимуществ. Экспоненциальное сглаживание предпочтительно выполнять с помощью пакетов программ для статистического анализа (STATISTICA, SPSS или, например, языка R).

Несколько большая подготовка понадобится для прогнозирования на базе авторегрессионных моделей или ИНС. Эти подходы подробно описаны в литературе, но свобода выбора характеристик таких моделей достаточно широка. Строгие формальные алгоритмы для этого не разработаны, поэтому качество проекций отчасти зависит от квалификации и интуиции специалиста, выполняющего прогноз. Методология АРПСС и ИНС реализована в пакетах статистических программ, ИНС также поддерживают большинство СКМ. Сложнее применить на практике байесовские сети или калмановскую фильтрацию. Многообразие вариантов моделей тут еще шире, а опыт их использования в области эпидемиологии сравнительно мал. Поэтому их употребление может потребовать дополнительного исследования. СКМ подходят для построения моделей в пространстве состояний и фильтров, а для создания байесовских сетей лучше применять специализированное программное обеспечение (например, WinBUGS).

Наиболее трудозатратным является использование МРЗ, так как они тесно связаны с задачей, для которой были разработаны. Для прогнозирования на основе эмпирических данных проще адаптировать классические аналитические модели, но степень их детализации позволяет достичь только грубых оценок заболеваемости. Высокодетализированные же МРЗ сложны в адаптации и настройке параметров. Выбор программного обеспечения для МРЗ, основанных на системах уравнений, достаточно велик — подойдут как СКМ, так и специальные средства моделирования динамических систем (Rand Model Designer и др.). Создание имитационной МРЗ для вычисления проекций, вероятно, придется выполнять вообще с нуля. Процедуры продвижения времени в агентных МРЗ сложнее, чем в прочих, поэтому предпочтительно пользоваться соответствующими средами моделирования (AnyLogic, NetLogo и проч.). Другие имитационные МРЗ можно реализовать и на базе ЯП.

Увы, универсального способа прогнозирования заболеваемости не существует — оптимальный подход следует выбирать, сравнивая результаты, полученные с помощью различных техник на основе эмпирических данных. Зачастую сложно отдать предпочтение одному методу прогнозирования — несколько подходов дают результаты сопоставимого качества. Для того чтобы сформировать рациональный прогноз из множества конкурирующих, целесообразно воспользоваться методами теории принятия решений, в частности многокритериального анализа [Рыков, Хорошилов, Щипин, 2005].

Список литературы

- Бейли Н. Математика в биологии и медицине. — М.: Мир, 1970. — 327 с.
- Белецкая С. Ю., Коровин В. Н., Родионов О. В. Разработка прогностических моделей развития заболеваемости детей в городском административном районе на основе нейросетевых технологий // Вестник Воронежского государственного технического университета. — 2010. — Т. 6, № 12. — С. 201–205.
- Боев Б. В., Макаров В. В. Компьютерное моделирование и прогнозирование эпидемий птичьего гриппа // Ветеринарная патология. — 2005. — № 3. — С. 49–58.
- Бокс Д., Дженкинс Г. Анализ временных рядов: прогноз и управление. Выпуск 1. — М.: Мир, 1974. — 406 с.
- Бородулин А. И., Десятков Б. М., Шабанов А. Н., Ярыгин А. А. Статистическая модель эпидемического процесса // Сибирский журнал индустриальной математики. — 2007. — Т. X, № 2 (30). — С. 23–30.
- Вьюн В. И., Еременко Т. К., Кузьменко Г. Е., Михненко Ю. А. Об одном подходе к прогнозированию эпидемиологической обстановки по гриппу-ОРВИ с использованием временных рядов // Математические машины и системы. — 2011. — Т. 1, № 2. — С. 131–136.
- Гальченко В. Я., Попов К. Р., Приземина И. Н., Качур Н. В. Прогнозирование временных рядов в задаче оценки эпидемической ситуации заболеваемости ОРВИ и гриппом по данным Луганской области // Український медичний альманах. — 2010. — Т. 13, № 2. — С. 20–22.
- Головин С. В., Петросян С. Л. Прогнозирование и моделирование развития заболеваемости в Воронежской области // Вестник Воронежского государственного технического университета. — 2009. — Т. 5, № 3. — С. 4–8.
- Ефимова Н. В., Горнов А. Ю., Зароднюк Т. С. Опыт использования искусственных нейронных сетей при прогнозировании заболеваемости населения (на примере г. Братска) // Экология человека. — 2010. — № 3. — С. 3–7.
- Иванов А. Г., Герасимова Л. И., Шувалова Н. В., Денисова Т. Г. Прогнозирование уровня заболеваемости и смертности от болезней системы кровообращения на региональном уровне на основе его многофазного математического моделирования // Медицинский альманах. — 2012. — № 3. — С. 20–24.

- Кизим Н. А., Доровской А. В. Анализ и прогнозирование тенденций заболеваемости населения Украины и Харьковской области // Проблемы экономики. — 2010. — № 3. — С. 39–44.
- Киселев О. И., Маринич И. Г., Карнова Л. С., Ежлова Е. Б., Лазикова Г. Ф., Ватолина А. А. Методика расчета эпидемических порогов по гриппу и острым респираторным вирусным инфекциям по субъектам Российской Федерации. — М.: НИИ гриппа Северо-Западного отделения РАМН, 2010. — 88 с.
- Научно-исследовательский институт гриппа. Система надзора за гриппом. — 2010. — URL: <http://www.influenza.spb.ru/system/> (дата обращения: 11.08.2013).
- Рыков А. С., Хорошилов В. О., Щупин К. С. Система прогнозирования инфекционной заболеваемости на основе многокритериального анализа временных рядов // Проблемы управления. — 2005. — № 1. — С. 26–32.
- Фирсов О. В. Гибридное прогнозирование заболеваемости раком почки и смертности от него на основе нейросетевых и статистических технологий // Врач-аспирант. — 2006. — № 1. — С. 15–32.
- Честнова Т. В., Смольянинова О. Л., Логвинов С. И. К вопросу о выборе метода математического анализа с целью прогнозирования заболеваемости лептоспирозом // Вестник новых медицинских технологий. — 2011. — Т. 18, № 4. — С. 18–21.
- A practical guide for designing and conducting influenza disease burden studies. — World Health Organization, 2008. — 49 p. — URL: http://www.wpro.who.int/emerging_diseases/documents/GuideforDesigningandConductingInfluenzaStudies/en/index.html (дата обращения: 13.08.2013).
- Abeku T. A., de Vlas S. J., Borsboom G., Teklehaimanot A., Kebede A., Olana D., van Oortmarssen G. J., Habbema J. D. F. Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best // Tropical Medicine and International Health. — 2002. — Vol. 7, № 10. — P. 851–857.
- Araz O. M., Fowler J. W., Lant T. W., Jehn M. A pandemic influenza simulation model for preparedness planning // Proceedings of the 2009 Winter Simulation Conference. — 2009. — P. 1986–1995.
- Bai Y., Jin Z. Prediction of SARS epidemic by BP neural networks with online prediction strategy // Chaos, Solitons and Fractals. — 2005. — Vol. 26, № 2. — P. 559–569.
- Bao L. A new infectious disease model for estimating and projecting HIV/AIDS epidemics // Sexually Transmitted Infections. — 2012. — Vol. 88, Suppl. 2. — P. i58–i64.
- Boyle J. R., Sparks R. S., Keijzers G. B., Crilly J. L., Lind J. F., Ryan L. M. Prediction and surveillance of influenza epidemics // Medical Journal of Australia. — 2011. — Vol. 194, № 4. — P. S28–S33.
- Brillman J. C., Burr T., Forslund D., Joyce E., Picard R., Umland E. Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance // BMC Medical Informatics and Decision Making. — 2005. — Vol. 5, № 4.
- Burkom H. S. Development, Adaptation, and Assessment of Alerting Algorithms for Biosurveillance // Johns Hopkins APL Technical Digest. — 2003. — Vol. 24, № 4. — P. 335–342.
- Burkom H. S., Murphy S. P., Shmueli G. Automated Time Series Forecasting for Biosurveillance // Statistics in Medicine. — 2007. — Vol. 26, № 22. — P. 4202–4218.
- Chao D. L., Halloran M. E., Obenchain V. J., Longini I. M., Jr. FluTE, a Publicly Available Stochastic Influenza Epidemic Simulation Model // PLoS Computational Biology. — 2010. — Vol. 6, № 1.
- Chatfield C., Yar M. Holt-Winters Forecasting: Some Practical Issues // Journal of the Royal Statistical Society. Series D (The Statistician). — 1988. — Vol. 37, № 2. — P. 129–140.
- Chen C. F., Ho W. H., Chou H. Y., Yang S. M., Chen I. T., Shi H. Y. Long-Term Prediction of Emergency Department Revenue and Visitor Volume Using Autoregressive Integrated Moving Average Model // Computational and Mathematical Methods in Medicine. — 2011. — Vol. 2011.
- Choi K., Thacker S. B. Mortality during Influenza Epidemics in the United States, 1967–1978 // American Journal of Public Health. — 1982. — Vol. 72, № 11. — P. 1280–1283.

- Coburn B. J., Wagner B. G., Blower S.* Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1) // *BMC Medicine*. — 2009. — Vol. 7, № 30.
- Colizza V., Barrat A., Barthélemy M., Vespignani A.* Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study // *BMC Medicine*. — 2007. — Vol. 5, № 34.
- Cowling B. J., Wong I. O. L., Ho L. M., Riley S., Leung G. M.* Methods for monitoring influenza surveillance data // *International Journal of Epidemiology*. — 2006. — Vol. 35, № 5. — P. 1314–1321.
- Daley D. J., Gani J.* Epidemic modelling: An introduction. — Cambridge University Press, 1999. — 225 p.
- Das T. K., Savachkin A. A., Zhu Y.* A large-scale simulation model of pandemic influenza outbreaks for development of dynamic mitigation strategies // *IIE Transactions*. — 2008. — Vol. 40, № 9. — P. 893–905.
- Díaz-Hierro J., Martín J. J. M., Arenas Á. V., González M. P. L. D., Arévalo J. M. P., González C. V.* Evaluation of time-series models for forecasting demand for emergency health care services // *Emergencias*. — 2012. — Vol. 24, № 3. — P. 181–188.
- Dugas A. F., Jalalpour M., Gel Y., Levin S., Torcaso F., Igusa T., Rothman R. E.* Influenza Forecasting with Google Flu Trends // *PLoS ONE*. — 2013. — Vol. 8, № 2.
- Eichner M., Schwehm M., Duerr H. P., Brockmann S. O.* The influenza pandemic preparedness planning tool Influsim // *BMC Infectious Diseases*. — 2007. — Vol. 7, № 17.
- Gardner E. S.* Exponential smoothing: The state of the art // *Journal of Forecasting*. — 1985. — Vol. 4, № 1. — P. 1–28.
- Hamilton J. D.* Time Series Analysis. — Princeton, NJ: Princeton University Press, 1994. — 820 p.
- Hethcote H. W.* The mathematics of infectious diseases // *SIAM Review*. — 2000. — Vol. 42, № 4. — P. 599–653.
- Hyndman R. J., Athanasopoulos G.* Forecasting: principles and practice. — 2012. — URL: <http://otexts.com/fpp/> (дата обращения: 28.08.2013).
- Hyndman R. J., Khandakar Y.* Automatic Time Series Forecasting: The forecast Package for R // *Journal of Statistical Software*. — 2008. — Vol. 27, № 3.
- Kiang R., Adimi F., Soika V., Nigro J., Singhasivanon P., Sirichaisinthop J., Leemingsawat S., Apiwathnasorn C., Looareesuwan S.* Meteorological, environmental remote sensing and neural network analysis of the epidemiology of malaria transmission in Thailand // *Geospatial Health*. — 2006. — Vol. 1, № 1. — P. 71–84.
- Lai D.* Monitoring the SARS Epidemic in China: A Time Series Analysis // *Journal of Data Science*. — 2005. — Vol. 3, № 3. — P. 279–293.
- Le Strat Y., Carrat F.* Monitoring epidemiologic surveillance data using hidden Markov models // *Statistics in Medicine*. — 1999. — Vol. 18, № 24. — P. 3463–3478.
- Li X., Tian H., Lai D., Zhang Z.* Validation of the Gravity Model in Predicting the Global Spread of Influenza // *International Journal of Environmental Research and Public Health*. — 2011. — Vol. 8, № 8. — P. 3134–3143.
- Moriña D., Puig P., Ríos J., Vilella A., Trilla A.* A statistical model for hospital admissions caused by seasonal diseases // *Statistics in Medicine*. — 2011. — Vol. 30, № 26. — P. 3125–3136.
- Myers M. F., Rogers D. J., Cox J., Flahault A., Hay S. I.* Forecasting Disease Risk for Increased Epidemic Preparedness in Public Health // *Advances in Parasitology*. — 2000. — Vol. 47. — P. 309–330.
- Ohkusa Y., Sugawara T.* Simulation model of pandemic influenza in the whole of Japan // *Japanese Journal of Infectious Diseases*. — 2009. — Vol. 62, № 2. — P. 98–106.

- Patel R., Longini I.M. Jr., Halloran M.E.* Finding optimal vaccination strategies for pandemic influenza using genetic algorithms // *Journal of Theoretical Biology.* — 2005. — Vol. 234, № 2. — P. 201–212.
- Pelat C., Boëlle P.Y., Cowling B.J., Carrat F., Flahault A., Ansart S., Valleron A.J.* Online detection and quantification of epidemics // *BMC Medical Informatics and Decision Making.* — 2007. — Vol. 7, № 29.
- Permanasari A.E., Rambli D.R.A., Dominic P.D.D.* Forecasting Method Selection Using ANOVA and Duncan Multiple Range Tests on Time Series Dataset // *Proceedings of the 2010 International Symposium on Information Technology (ITSim 2010).* — 2010. — Vol. 2. — P. 941–945.
- Reis B.Y., Mandl K.D.* Time Series Modeling for Syndromic Surveillance // *BMC Medical Informatics and Decision Making.* — 2003. — Vol. 3, № 2.
- Rinaldo A., Bertuzzo E., Mari L., Righetto L., Blokesch M., Gatto M., Casagrandi R., Murray M., Vesenbeckh S.M., Rodriguez-Iturbe I.* Reassessment of the 2010–2011 Haiti cholera outbreak and rainfall-driven multiseason projections // *Proceedings of the National Academy of Sciences of the United States of America.* — 2012. — Vol. 109, № 17. — P. 6602–6607.
- Saramäki J., Kaski K.* Modelling development of epidemics with dynamic small-world networks // *Journal of Theoretical Biology.* — 2005. — Vol. 234, № 3. — P. 413–421.
- Schmidt R., Waligora T.* Influenza Forecast: Case-Based Reasoning or Statistics? // *Proceedings of the 11th international conference on Knowledge-Based Intelligent Information and Engineering Systems: Part I. Series: Lecture Notes in Computer Science.* — 2007. — Vol. 4692. — P. 287–294.
- Sebastiani P., Mandl K.D., Szolovits P., Kohane I.S., Ramoni M.F.* A Bayesian dynamic model for influenza surveillance // *Statistics in Medicine.* — 2006. — Vol. 25, № 11. — P. 1803–1816.
- Serfling R.E.* Methods for Current Statistical Analysis of Excess Pneumonia-influenza Deaths // *Public Health Reports.* — 1963. — Vol. 78, № 6. — P. 494–506.
- Shaman J., Karspeck A.* Forecasting seasonal outbreaks of influenza // *Proceedings of the National Academy of Sciences of the United States of America.* — 2012. — Vol. 109, № 50. — P. 20425–20430.
- Shmueli G., Fienberg S.E.* Current and Potential Statistical Methods for Monitoring Multiple Data Streams for Biosurveillance // *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication.* — New York: Springer Science + Business Media, 2006. — P. 109–140.
- Siettos C.I., Russo L.* Mathematical modeling of infectious disease dynamics // *Virulence.* — 2013. — Vol. 4, № 4. — P. 1–12.
- Sirakoulis G.Ch., Karafyllidis I., Thanailakis A.* A cellular automaton model for the effects of population movement and vaccination on epidemic propagation // *Ecological Modelling.* — 2000. — Vol. 133, № 3. — P. 209–223.
- Soebiyanto R.P., Adimi F., Kiang R.K.* Modeling and Predicting Seasonal Influenza Transmission in Warm Regions Using Climatological Parameters // *PLoS ONE.* — 2010. — Vol. 5, № 3.
- Soebiyanto R.P., Kiang R.K.* Modeling Influenza Transmission Using Environmental Parameters // *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science.* — 2010. — Vol. XXXVIII, part 8. — P. 330–334.
- Sumi A., Kamo K.* MEM spectral analysis for predicting influenza epidemics in Japan // *Environmental Health and Preventive Medicine.* — 2012. — Vol. 17, № 2. — P. 98–108.
- Thompson W.W., Shay D.K., Weintraub E., Brammer L., Cox N., Anderson L.J., Fukuda K.* Mortality Associated With Influenza and Respiratory Syncytial Virus in the United States // *The Journal of the American Medical Association.* — 2003. — Vol. 289, № 2. — P. 179–186.

- Unkel S., Farrington C. P., Garthwaite P. H., Robertson C., Andrews N.* Statistical methods for the prospective detection of infectious disease outbreaks: a review // *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. — 2012. — Vol. 175, № 1. — P. 49–82.
- Vergu E., Grais R., Sarter H., Fagot J. P., Lambert B., Valleron A. J., Flahault A.* Medication Sales and Syndromic Surveillance, France // *Emerging Infectious Diseases*. — 2006. — Vol. 12, № 3. — P. 416–421.
- Viboud C., Boëlle P. Y., Carrat F., Valleron A. J., Flahault A.* Prediction of the Spread of Influenza Epidemics by the Method of Analogues // *American Journal of Epidemiology*. — 2003. — Vol. 158, № 10. — P. 996–1006.
- Watkins R. E., Eagleson S., Veenendaal B., Wright G., Plant A. J.* Disease surveillance using a hidden Markov model // *BMC Medical Informatics and Decision Making*. — 2009. — Vol. 9, № 39.